

CLAIMS

1. A speech recognition dictionary creation device that creates a speech recognition dictionary, said device comprising:

an abbreviated word generation unit operable to generate an abbreviated word of a recognition object that is made up of one or more constituent words based on a rule that takes into account ease of pronunciation; and

a vocabulary storage unit operable to hold, as the speech recognition dictionary, the generated abbreviated word together with the recognition object.

2. The speech recognition dictionary creation device according to Claim 1, further comprising:

a word division unit operable to divide the recognition object into the constituent words; and

a mora string generation unit operable to generate mora strings of the respective constituent words based on readings of the respective divided constituent words,

wherein said abbreviated word generation unit is operable to generate the abbreviated word made up of one or more moras by extracting one or more moras from the mora strings of the respective constituent words and concatenating the extracted moras based on the mora strings of the respective constituent words generated by said mora string generation unit.

3. The speech recognition dictionary creation device according to Claim 2,

wherein said abbreviated word generation unit includes:

an abbreviated word generation rule storage unit operable to hold a generation rule for generating an abbreviated word using moras;

a candidate generation unit operable to generate candidate

abbreviated words, each being made up of one or more moras, by extracting one or more moras from the mora strings of the respective constituent words and concatenating the extracted moras; and

5 an abbreviated word determination unit operable to determine an abbreviated word for final generation, by applying the generation rule held by said abbreviated word generation rule storage unit to the generated candidate abbreviated words.

10 4. The speech recognition dictionary creation device according to Claim 3,

 wherein said abbreviated word generation rule storage unit is operable to hold a plurality of generation rules,

 said abbreviated word determination unit is operable to
15 calculate a likelihood under each of the generation rules stored in said abbreviated word generation rule storage unit and to determine an utterance probability by comprehensively taking into account the calculated likelihoods, the utterance probability being determined for each of the generated candidate abbreviated words, and

20 said vocabulary storage unit is operable to hold the abbreviated word and the utterance probability that are determined by said abbreviated word determination unit.

25 5. The speech recognition dictionary creation device according to Claim 4,

 wherein said abbreviated word determination unit is operable to determine the utterance probability by summing up values that are obtained by multiplying the likelihoods for the respective generation rules by corresponding weighting factors.

30

6. The speech recognition dictionary creation device according to Claim 5,

wherein said abbreviated word determination unit is operable to determine that a candidate abbreviated word is the abbreviated word for final generation in the case where the utterance probability of the candidate abbreviated word exceeds a predetermined threshold.

7. The speech recognition dictionary creation device according to Claim 4,

wherein said abbreviated word generation rule storage unit is operable to hold a first rule concerning dependency relationship between words, and

said abbreviated word determination unit is operable to determine, based on the first rule, the abbreviated word for final generation from among the candidates.

8. The speech recognition dictionary creation device according to Claim 7,

wherein the first rule includes a condition that an abbreviated word should be generated using a modifier and a modified word as a pair.

9. The speech recognition dictionary creation device according to Claim 7,

wherein the first rule includes a rule indicating a relationship between the likelihood and a distance between a modifier and a modified word that make up an abbreviated word.

10. The speech recognition dictionary creation device according to Claim 4,

wherein said abbreviated word generation rule storage unit is operable to hold a second rule that is related to at least one of a length of a partial mora string and a position of the partial mora

string, the length being a length of the partial mora string that is extracted from a mora string of the constituent word when an abbreviated word is generated, and the position being a position of the partial mora string in the constituent word, and

5 said abbreviated word determination unit is operable to determine, based on the second rule, the abbreviated word for final generation from among the candidates.

11. The speech recognition dictionary creation device according
10 to Claim 10,

 wherein the second rule includes a rule indicating a relationship between the likelihood and a number of moras indicating the length of the partial mora string.

12. The speech recognition dictionary creation device according
15 to Claim 10,

 wherein the second rule includes a rule indicating a relationship between the likelihood and a number of moras indicating a distance from a top of the constituent word to the partial
20 mora string, the distance indicating the position of the partial mora string in the constituent word.

13. The speech recognition dictionary creation device according
 to Claim 4,

25 wherein said abbreviated word generation rule storage unit is operable to hold a third rule related to concatenated partial mora strings that make up an abbreviated word, and

 said abbreviated word determination unit is operable to determine, based on the third rule, the abbreviated word for final
30 generation from among the candidates.

14. The speech recognition dictionary creation device according

to Claim 13,

wherein the third rule includes a rule indicating a relationship between the likelihood and a combination of a last mora and a top mora, the last mora being included in a former of the concatenated two partial mora strings and the top mora being included in a latter of the concatenated two partial mora strings.

15. The speech recognition dictionary creation device according to Claim 2, further comprising:

10 an extraction condition storage unit operable to hold a condition for extracting the recognition object from character string information that includes the recognition object;

a character string information obtainment unit operable to obtain the character string information that includes the recognition object; and

15 a recognition object extraction unit operable to extract the recognition object from the character string information obtained by said character string information obtainment unit according to the condition held by said extraction condition storage unit, and to send the extracted recognition object to said word division unit.

16. A speech recognition device that recognizes an input speech by comparing the input speech with a model corresponding to a vocabulary registered in a speech recognition dictionary, said device comprising

25 a recognition unit operable to recognize the speech using the speech recognition dictionary created by the speech recognition dictionary creation device according to Claim 1.

30 17. The speech recognition device according to Claim 16, wherein the abbreviated word and the utterance probability of the abbreviated word are registered into the speech recognition

dictionary together with the recognition object, and

said recognition unit is operable to recognize the speech by taking into account the utterance probability registered in the speech recognition dictionary.

5

18. The speech recognition device according to Claim 17,

wherein said recognition unit is operable (i) to generate a candidate for a recognition result of the speech and a likelihood of the candidate, (ii) to add a likelihood corresponding to the utterance probability to the generated likelihood, and (iii) to output the candidate as a final recognition result based on the resulting addition value.

10

19. The speech recognition device according to Claim 16, further comprising:

15

an abbreviated word use history storage unit operable to hold, as use history information, an abbreviated word recognized for the speech and a recognition object corresponding to the abbreviated word; and

20

an abbreviated word generation control unit operable to control generation of an abbreviated word by the abbreviated word generation unit based on the use history information held by said abbreviated word use history storage unit.

25

20. The speech recognition device according to Claim 19,

wherein the abbreviated word generation unit of the speech recognition dictionary creation device includes:

an abbreviated word generation rule storage unit operable to hold a generation rule for generating an abbreviated word using moras;

30

a candidate generation unit operable to generate candidate abbreviated words, each being made up of one or more moras, by

extracting one or more moras from the mora strings of the respective constituent words and concatenating the extracted moras; and

an abbreviated word determination unit operable to
5 determine an abbreviated word for final generation, by applying the generation rule held by said abbreviated word generation rule storage unit to the generated candidate abbreviated word, and

said abbreviated word generation control unit is operable to control the generation of the abbreviated word by making one of
10 change, deletion, and addition to the generation rule held by the abbreviated word generation rule storage unit.

21. The speech recognition device according to Claim 16, further comprising:

15 an abbreviated word use history storage unit operable to hold, as use history information, an abbreviated word recognized for the speech and a recognition object corresponding to the abbreviated word; and

a dictionary revision unit operable to revise the abbreviated
20 word stored in the speech recognition dictionary based on the use history information held by said abbreviated word use history storage unit.

22. The speech recognition device according to Claim 21,

25 wherein the abbreviated word and the utterance probability of the abbreviated word are registered into the speech recognition dictionary together with the recognition object, and

said dictionary update unit is operable to revise the abbreviated word by changing the utterance probability of the
30 abbreviated word.

23. A speech recognition device that recognizes an input speech

by comparing the input speech with a model corresponding to a vocabulary registered in a speech recognition dictionary, said device comprising:

- the speech recognition dictionary creation device according to
- 5 Claim 1; and
- a recognition unit operable to recognize the speech using the speech recognition dictionary created by said speech recognition dictionary creation device.

10 24. A speech recognition dictionary creation method for creating a speech recognition dictionary, said method comprising the steps of:

- generating an abbreviated word of a recognition object that is made up of one or more constituent words based on a rule that takes
- 15 into account ease of pronunciation; and
- registering, into the speech recognition dictionary, the generated abbreviated word together with the recognition object.

20 25. The speech recognition dictionary creation method according to Claim 24, further comprising:

- dividing the recognition object into the constituent words;
- and
- generating mora strings of the respective constituent words based on readings of the respective divided constituent words,
- 25 wherein in said generating of the abbreviated word, the abbreviated word made up of one or more moras is generated by extracting one or more moras from the mora strings of the respective constituent words and concatenating the extracted moras based on the mora strings of the respective constituent words
- 30 generated by said mora string generation unit.

26. A speech recognition method for recognizing an input speech

by comparing the input speech with a model corresponding to a vocabulary registered in a speech recognition dictionary, said method comprising the step of

5 recognizing the speech using the speech recognition dictionary created by the speech recognition dictionary creation method according to Claim 24.

27. A speech recognition method for recognizing an input speech by comparing the input speech with a model corresponding to a
10 vocabulary registered in a speech recognition dictionary, said method comprising:

the steps included in the speech recognition dictionary creation method according to Claim 24; and

15 a step of recognizing the speech using the speech recognition dictionary created by the speech recognition dictionary creation method.

28. A program for a speech recognition dictionary creation device that creates a speech recognition dictionary, said program causing a
20 computer to execute the steps included in the speech recognition dictionary creation method according to Claim 24.

29. A program for a speech recognition device that recognizes an input speech by comparing the input speech with a model
25 corresponding to a vocabulary registered in a speech recognition dictionary, said program causing a computer to execute the step included in the speech recognition method according to Claim 26.